

Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science* e *Big Data*

Dalton Luiz Schneckenberg Filho

Modelo de Regressão de Cox

**Curitiba
2020**

Dalton Luiz Schneckenberg Filho

Modelo de Regressão de Cox

Monografia apresentada ao Programa de Especialização em *Data Science* e *Big Data* da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. José Luiz Padilha

Curitiba
2020

Modelo de Regressão de Cox - Uma Aplicação a Dados de Pacientes com a Doença de Chagas

Cox Regression Model - An Application to Data from Patients with Chagas Disease

Dalton Luiz S. Filho¹

¹Departamento de Estatística, Universidade Federal do Paraná Av. Cel. Francisco H. dos Santos, 100 - Jardim das Américas, Curitiba - PR, 81530-000, Brasil*

A análise de sobrevivência é uma das áreas da estatística que mais cresceu nas últimas duas décadas. A razão deste crescimento é o desenvolvimento e aprimoramento de técnicas estatísticas combinadas com computadores cada vez mais velozes. Este trabalho tem por objetivo utilizar conceitos do modelo semi-paramétrico de sobrevivência, em especial o modelo de Cox, para modelar o tempo de vida de doentes portadores da doença de chagas. O modelo de Cox estima a função de risco e muitas vezes é interessante estimar o efeito das covariáveis. Desta forma Cox propôs o modelo semi-paramétrico de riscos proporcionais que permite a análise de dados resultantes de estudos do tempo de vida, onde a variável resposta é o tempo de ocorrência de um evento de interesse ajustado por covariáveis.

Palavras-chave: Modelos semi-paramétricos. Modelo de Cox. Análise de sobrevivência

Survival analysis is one of the fastest growing areas of statistics in the past two decades. The reason for this growth is the development and improvement of statistical techniques combined with increasingly fast computers. This work aims to use concepts of the semi-parametric model of survival, especially the Cox model, to model the life span of patients with Chagas disease. Cox's model estimates the risk function and it is often interesting to estimate the effect of covariates. In this way, Cox proposed the semi-parametric model of proportional risks that allows the analysis of data resulting from life time studies, where the response variable is the time of occurrence of an event of interest adjusted by covariates.

Keywords: Semi-parametric models. Cox model. Survival analysis

1. Introdução

A doença de Chagas (ou Tripanossomíase americana) é a infecção causada pelo protozoário *Trypanosoma cruzi*. Apresenta uma fase aguda (doença de Chagas aguda – DCA) que pode ser sintomática ou não, e uma fase crônica, que pode se manifestar nas formas indeterminada, cardíaca, digestiva ou cardiodigestiva. Em função das ações de controle de vetores realizadas a partir da década de 1970, o Brasil recebeu em 2006 a certificação Internacional da interrupção da transmissão vetorial pelo *Triatoma infestans*, espécie exótica e responsável pela maior parte da transmissão vetorial no passado. Porém, estima-se que existam aproximadamente 12 milhões de portadores da doença crônica nas Américas, e que haja no Brasil, atualmente, pelo menos um milhão de pessoas infectadas por *T. cruzi*.

De 2008 a 2017 foram registrados no Brasil um total de 2172 casos agudos da doença de chagas, sendo a região norte (Pará) onde foi verificada a maior incidência. A taxa média de mortalidade registrada no Brasil de 2015 a 2017 foi de 2,19 para cada 100.000 habitantes.

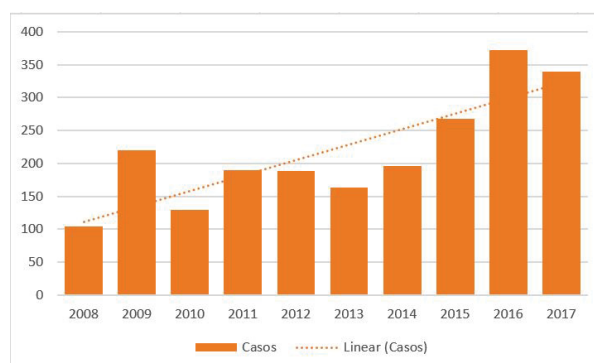


Figura 1: Casos da doença de chagas registrados no Brasil de 2008 a 2017.

*daltonluiz26@gmail.com

New York Heart Association) trata-se da classificação do nível de severidade do problema cardíaco do paciente, podendo variar do nível 1 (doença cardíaca, sem limitações para atividade física) ao 4 (doença cardíaca que acarreta em incapacidade para exercer qualquer atividade física).

Os resultados corroboraram a citação da pesquisadora.

Para os níveis 2, 3 e 4 verificamos o risco crescente em 24, 37 e 66% respectivamente em relação à referência, resultados realmente expressivos.

A segunda variável avaliada foi aneurismaVD, trata-se de uma variável dicotômica (sim/não) que representa a presença de aneurisma no ventrículo direito do paciente, fato que se não verificado contribui com a redução de 38% no risco de morte.

O índice de concordância apresentado na análise foi de 0,82.

A figura 3 traz a análise completa dos dados para referência.

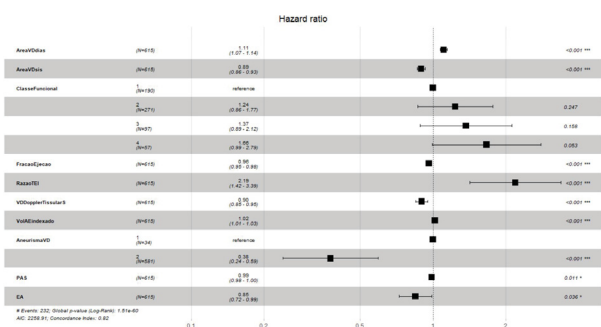


Figura 3: Análise de riscos, HR > 1 indica risco aumentado de morte

4.2. Validação do modelo

Embora o modelo tenha sido ajustado, ainda é necessária a verificação do atendimento da suposição dos riscos proporcionais para que o modelo de Cox seja considerado adequado.

De acordo com Colosimo e Giolo (2006), técnicas gráficas aplicadas aos diferentes resíduos obtidos na regressão são particularmente muito utilizadas para examinar a distribuição dos erros. Estas técnicas, como observados em Klein e Moeschberger (1997), devem ser utilizadas primariamente para rejeitar modelos inadequados e não para “provar” que um determinado modelo é correto.

Os métodos gráficos aplicados ao modelo em questão utilizaram o logaritmo da função de risco acumu-

lada e os resíduos padronizados de Schoenfeld. Ambos devem ser construídos para cada variável do modelo.

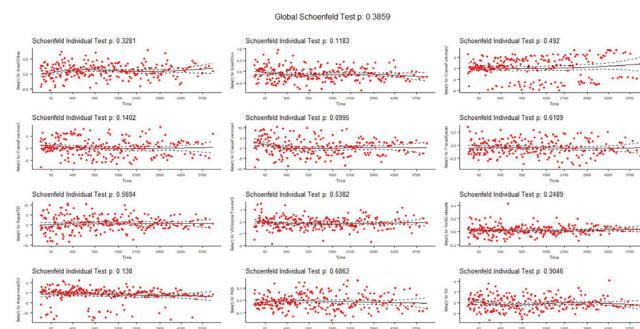


Figura 4: O Padrão aleatório verificado nos resíduos Schoenfeld não rejeita o modelo

5. Conclusão

No presente trabalho foi estudado o modelo de cox e sua utilização na análise de sobrevivência para explicação dos efeitos das variáveis de controle na ocorrência do evento de interesse.

Agradecimentos

Agradeço ao Banco Bradesco pela confiança e investimento no desenvolvimento contínuo dos funcionários. À minha esposa e filhos pelo apoio incondicional e compreensão nos períodos de ausência, à coordenação do Curso de Data Science Big Data da UFPR, seu corpo docente e administração que oportunizaram intenso avanço no conhecimento, e finalmente ao meu orientador José Luiz Padilha Silva, pela paciência e auxílio despendidos no desenvolvimento do presente trabalho.

Referências

- [1] Ministério da Saúde do Brasil, <https://saude.gov.br/saude-de-a-z/doenca-de-chagas>
- [2] Análise de sobrevivência aplicada, Enrico Antônio Colosimo, Sueli Ruiz Giolo. - São Paulo: Blucher, 2006.
- [3] Multivariate Imputation by Chained Equations in R, Stef VAN BUUREN and Karin Groothuis-Oudshoorn, Journal of Statistical Software, 2011, <https://www.jstatsoft.org/v45/i03/>
- [4] Universidade Federal de Minas Gerais, http://www.est.ufmg.br/enricoc/pdf/avancados_medicinal/aula12_15_cox.pdf